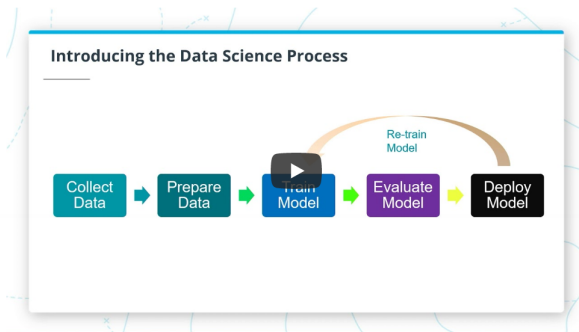


The Data Science Process

Big data has become part of the lexicon of organizations worldwide, as more and more organizations look to leverage data to drive informed business decisions. With this evolution in business decision-making, the amount of raw data collected, along with the number and diversity of data sources, is growing at an astounding rate. This data presents enormous potential.

Raw data, however, is often noisy and unreliable and may contain missing values and outliers. Using such data for modeling can produce misleading results. For the data scientist, the ability to combine large, disparate data sets into a format more appropriate for analysis is an increasingly crucial skill.

The data science process typically starts with collecting and preparing the data before moving on to training, evaluating, and deploying a model. Let's have a look.



QUESTION 1 OF 3

Here are the typical steps of the data science process that we just discussed. Can you remember the correct order?

Submit to check your answer choices

STEP	DESCRIPTION
Steps 1 & 2	Collect and prepare the data
Steps 3 & 4	Train the model and evaluate its performance
Steps 5 & 6	Deploy the model and then retrain as necessary

SUBMIT

QUESTION 2 OF 3

Here are some of the steps once again, along with some of the actions that you would carry out during those steps. Can you match the step with the appropriate action?

(Again, first try to do it from memory—but have a look at the text or video above if you get stuck.)

Submit to check your answer choices

ACTION	WHICH STEP OF THE PROCESS?
Package the model and dependencies	Deploy the model
Run the model through a final exam using data from your validation data set	Evaluate the model
Create features needed for the model	Prepare the data
Select the algorithm, and prepare training, testing, and validation data sets	Train the model

SUBMIT

QUESTION 3 OF 3

In machine learning, often you have to tune parameters for the chosen learning algorithm to improve the performance on relevant metrics, such as prediction accuracy. At what stage of the data science lifecycle do you optimize the parameters?

- ☒ Training the model
- ☐ Evaluating the model
- ☐ Deploying the model

SUBMIT

NEXT